DOCUMENT RESUME

ED 366 225 FL 021 851

AUTHOR Aweiss, Salem
TITLE Comprehension Assessment Measures in Second and Foreign Language Reading.
PUB DATE 93
NOTE 31p.; Paper presented at the Annual Meeting of the Central States Conference on the Teaching of Foreign Languages (1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Evaluation Methods; *Language Tests; *Rating Scales; *Reading Comprehension; *Reading Tests; Second Language Instruction; *Second Languages

ABSTRACT

Assessment of second language reading comprehension has evolved from a relatively narrow conceptualization of reading as a process of mastering hierarchically ordered subskills, with the author as primary creator of meaning, to reading as an interaction among reader, author, and text. Reading assessment has several purposes: sorting students; diagnosing individual problems; and evaluating instructional effectiveness. Most measurement methods are based on a psychometric perspective, but a cognitive approach, which sees reading as a constructive process, may give more insight into why a learner is able or unable to master specific objectives. Conventional measures of reading comprehension include multiple-choice questions, short-answer questions, and cloze tests. Currently, text recall is considered the best method for inferring comprehension. Criticism of the method focuses on its inappropriateness for the English-as-a-Second-Language situation, absence of an objective weighting and analyzing system, time consumed, holistic approach, and lack of differentiation of processes and skills used. However, a constructive activity scale can be used with a recall protocol to identify the cognitive activities involved in text comprehension. Such a scale would analyze activities on four levels: prepropositional/fragmented associations, knowledge/details retelling, assimilation, and problem-solving and integration. The proposed method allows for both quantitative and qualitative assessment. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 366 225

COMPREHENSION ASSESSMENT MEASURES IN SECOND AND FOREIGN LANGUAGE READING

SALEM AWEISS, Ph.D.

THE OHIO STATE UNIVERSITY

FL021851

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Aweiss

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

INTRODUCTION

Despite the rapid spread of oral media, the acquisition of literacy skills remains paramount for learners of second and foreign languages all over the world. Bernhardt (1991) notes that interest in second language literacy skills has grown in the last decade for social-political, pedagogical, and cognitive reasons.

Any discussion about reading comprehension would not be complete without a serious consideration of comprehension assessment. At present, reading comprehension assessment is probably one of the most, if not the most important field in second language reading research. Unless reliable and valid measures of comprehension are developed, it is safe to argue along with Bernhardt (1991) that "the entire area of reading research will remain uncertain" (p. 233).

Assessment measures are tasks to be observed to gain information. They are samples of behavior. Information gleaned should be viewed as an integral part of the instructional process that informs and empowers students and instructors. It is thus obvious that assessment is multidimensional. It is a necessary component of any type of effective instruction, helping us to answer many questions. Comprehension assessment that seeks to support instructional decision making must consider how the various facets of reading--text-driven and knowledge-driven--may be affecting comprehension performance. Bernhardt (1984) argues that generic tests that do not consider reader background knowledge may be biased and may therefore, be invalid because they fail to accurately measure reading ability.

Since our goal is to enable our students to comprehend a variety of texts independently, one of the most important questions we need to answer in comprehension assessment is how well they are achieving this goal. The answer to this question is very important, but it also leads to a further question, i.e.,"How can we help them comprehend better?" Assessment must not merely tell us about comprehension as a product. It must give us some insight into a reader's

comprehension process because a primary goal of assessment is to inform and guide instruction.

Thus an important focus of comprehension assessment should be to determine under what instructional conditions a learner comprehends best. Assessment that seeks to answer this question is referred to as dynamic assessment and is sometimes labeled as interactive (Campione and Brown, 1985; Valencia, Pearson, 1988; Wang, 1987; Wood, 1988). As instructors interact with students and texts and model strategic reading processes, they look for patterns in how students construct meaning. This procedure, in turn, informs and shapes decisions about materials, tasks, pacing, and feedback for future lessons.

Although literacy assessment is currently under scrutiny and reconceptualization and its techniques are undergoing change, reading programs have been slow or reluctant to examine traditional assessment methods that remain dominant through the United States and in many parts of the world that have been influenced by U.S. practices. Many reading programs rely exclusively on one standardized reading test not only to place and diagnose incoming students but also to evaluate program effectiveness (Wood, 1989).

To remedy the situation, reading professionals have continuously searched for better reading comprehension assessment measures. Increasingly, alternative approaches to assessment seek to attend to complex learning and processes. Unfortunately the task is not that simple. Reading specialists are faced with the enormous task of conceptualizing a comprehensive model of assessment that reflects current reading research and theory, and is appropriate to the philosophy and goals of their programs, and is unique to the characteristics of their students. Obviously, this task is not easy.

Reading Comprehension

Theoretical and empirical interest in reading comprehension is a rather recent

phenomenon. In fact, there was little systematic research on reading comprehension before the 1960's, with books of the time not even mentioning the phrase "reading comprehension," let alone giving it much treatment (e.g., Anderson & Deerborn, 1952; Woodworth, 1938). In the 1970s, things changed. Rothkopf (1972), for example, insisted that research on reading should emphasize comprehension and its most effective facilitation. More recently, Daneman & Tardif (1988) argue that to preclude comprehension from a study of reading would be to invite theories of reading that are incomplete and of no practical relevance.

Venezky (1984) notes that comprehension was not considered important by researchers at the beginning of this century because, to them, reading usually meant oral reading, and comprehension of a given text was simply assumed if a reader's "pronunciation was correct and natural" (p. 13). The importance of comprehension became more salient with the advent of the testing movement because of its interest in the assessment of comprehension ability. Greater interest in language comprehension, moreover, was partially, a result of developments in the disciplines of human factors research, computer science, and linguistics. These developments resulted in a shift in psychology from behaviorist to an information-processing or cognitive orientation. As a consequence, constructs from computer science (such as knowledge representation, buffers and working memory, and parallel, serial, and interactive processes) as well as constructs from linguistics (such as discourse structures, integration, and inferencing) have become part of the parcel of the theoretically oriented field of reading comprehension research (Ashcraft, 1989).

In sum, reading comprehension theory and research have made tremendous strides over the past three decades-from a narrow conceptualization of reading as a process of mastering a number of hierarchically ordered subskills and one that acknowledges the primary role of the author as the creator of meaning, to a broader,

more reader-based conceptualization of reading as the interaction among reader, author, and text. Notions about the residence of the meaning (with the author, with the text, with the interaction among reader, author and text) have changed over the period, as well as the kind of knowledge thought important in the act of reading (declarative or procedural).

Purposes of Reading Assessment

Cronbach (1984) defines tests as "systematic procedures for observing behavior and describing it with the aid of numerical scales or fixed categories" (Cronbach, 1984, p. 26). On the basis of information of the observed performance on a test, inferences are made about the more general underlying competence.

In the language context, Weir (1990) notes that in testing language ability, we are evaluating samples of performance in certain specific contexts of use, created under particular test constraints, for what they can tell us about a person's communicative capacity or language ability.

Cross and Paris (1987) discuss three major purposes for reading assessment: sorting, diagnosing, and evaluating. Reading tests are used to sort students by arranging them along a continuum from highest to lowest scores. Sorting is used to predict learners' academic success or to indicate mastery of an instructional program. This type of measurement also functions as a formative measure of assessment. It provides information that directs subsequent teaching-learning activity. Formative testing aids in decision-making on how instruction is to be shaped. It gives information which helps the instructor.

The second purpose of assessment, diagnosing individuals' reading problems, calls for gathering information about a particular student's strategies and processes. The diagnostic findings should be used to make informed decisions about individuals, not decisions about group changes. Diagnostic testing provides information about intrapersonal (within the person) factors that will influence the

teaching-learning process for that individual.

The final purpose of assessment, evaluating, calls for determining whether a particular experimental treatment or instructional program has had an effect on dependent variables such as improved reading performance.

An Historical Review:Psychometric vs.Cognitive Approaches to Assessment and Reading Comprehension

Despite the fact that psychologists, educators, and reading specialists have been concerned with research, evaluation, and training of reading comprehension for several decades (cf. Huey, 1908; Thorndike, 1917; Dewey, 1935; Davis, 1944, 1972; Thorndike, 1973; Johnston, 1983; among others), the measures and analyses of "reading comprehension" are still being debated.

Throughout the years, different assessment methods have served as comprehension measures. Recent research from two distantly related enterprises, cognitive sciences and research on teaching, has encouraged reading educators to rethink prevailing constructs about reading and how they affect reading comprehension.

Most of the current methods used are grounded in the psychometric paradigm of assessment. Recently, new ideas in comprehension testing have been advanced from information-processing and interactive learning perspectives. What distinguishes psychometric from cognitive approaches is this emphasis of psychometrics on products or factors, rather than processes and its emphasis on comparison and description of methods rather than on experimentation. Psychometric measures tell us if students master the designated instructional objectives by indicating whether they get an item right or wrong. Assessment procedures congruent with cognitive psychology may shed light on why learners are able or unable to master their designated instructional objectives. It is obvious that these assessment procedures go beyond the surface level of knowledge and assess how

deeply learners have organized knowledge or to what extent the students have linked concepts with other concepts.

A Cognitively Based View of Learning and Reading Comprehension

A central premise of cognitive psychology is that comprehension is a constructive process involving information from the environment and from semantic memory (Doyle 83). Humans respond to an external stimulus based on the stimulus itself and upon past experience retrieved from long-term memory which is relevant to the stimulus.

Reading is a far more complex process than had been envisioned by early reading researchers; above all, it is not a set of skills to be mastered. In the traditional view, readers are passive recipients of information in the text. Meaning resides in the text itself, and the goal of the reader is to reproduce that meaning.

Current research in text processing is predicated on the idea that comprehension is an active process of construction rather than simple information reception. Cognitive approaches to reading comprehension generally recognize that meaning does not rest with the text. Instead, they emphasize the interactive nature of reading (Rumelhart and Ortony, 1977; Bernhardt, 1991) and the constructive nature of comprehension (Anderson, Reynolds, Schallert, & Goetz, 1977; Rumelhart, 1980; Spiro, 1980, Bernhardt, 1991). Reading is a constructive process that combines individual units to form new configurations; that is, there is some type of cognitive constructive activity involved in the process of reading. This constructive activity according to Page (1990), is oriented towards the construction of a network of information or model which comprises all the textual information. Unless the reader is reading a text from a very specific perspective which is different from that of the author, the task he engages in when he reads a text seriously is the construction of a structured representation which resembles closely the structure which the author has given to the information deposited in his text.

Background Knowledge and Reading Comprehension

All readers, both novices and experts, use their existing knowledge and a range of cues from the text and the situational context in which reading occurs to build, or construct, a model of meaning from the text. The knowledge that readers bring to the text is paramount (Anderson, Reynolds, Shallert, & Goetz, 1977; Rumelhart & Ortony, 1977; Spiro, 1980; among others). Across all levels of age and ability, readers use their existing knowledge as a filter to interpret and construct meaning a given text (Anderson & Pearson, 1984). They also use this knowledge to determine importance (Afflerbach, 1990,1986), to draw inference (Fincher-Keifer, 1992; Hansen, 1981; Hansen & Pearson, 1983), to elaborate text (Hansen, and Pearson, 1983), and to monitor comprehension (Dewitz, Carr, Patberg, 1987; Casanave, 1988).

In sum, skilled readers use their stores of existing knowledge as well as a number of flexible strategies to construct a mental model of the text. They monitor their ongoing comprehension and change strategies when comprehension breaks down. They adjust their strategy selection and their metacognitive awareness depending on their level of domain-specific knowledge (Alexander & Judy, 1988).

Measures of Reading Comprehension

In order to better define the construct of reading comprehension, the focus in many recent studies into reading has moved away from product to investigating the reading process (Farr, Pritchard, & Smitten, 1990; Pritchard, 1990). The change in our thinking about how the printed word is understood, however, has not been accompanied by a change in our practices and the methods we use to measure that understanding. Cognitively based research suggests a reconceptualization of the reading process and, therefore, a reconceptualization of the comprehension curriculum and comprehension assessment.

Like reading comprehension, reading comprehension assessment is a complex process involving a variety of measures. These measures vary not only

according to the type of questions that they seek to answer but also according to their structure (Valencia, McGinley, & Pearson 1991). Comprehension assessment tools range from the unstructured and spontaneous gathering of information during instruction to structured tests with specifically defined outcomes and directions for administration and scoring. In the middle of the continuum are semi-structured measures, informal but planned assessments that require "more input and interpretation from the teacher and/or provide greater latitude in student response (Valencia, McGinley, & Pearson 1991).

Conventional Measures of Reading Comprehension Assessment

Multiple Choice: Although the multiple-choice test format is one of the most frequently used test formats (Anderson et al, 1992; Klein-Braley, 1984, 1990; Nivo, 1989), it has frequently been criticized because the correct answer can be reached in more than one way and can often be identified "without actually understanding the text and without any judgmental activity in selecting the correct response" (Klein-Braley, 1985, cited in Nivo, 1989). According to Klein-Braley (1984, 1985, 1991) the process of reaching the correct answer on reading comprehension test thus may not reflect the processes involved in actual reading.

Bernhardt (1991), Henning (1987), Pyrczak (1975) assert that multiple choice test items open the way to guessing and can often be answered without reference to the reading passage. Further, because of the difficulty involved in producing multiple choice questions that assess whether or not the student has been able to integrate passage information, this test mode often taps knowledge at the discrete-point level. Thus, the potential for assessing meaning that the reader has gleaned from the text is sacrificed.

Short Answer Questions: Weir (1990) argues that this technique is extremely useful for testing both reading and listening comprehension. This format allows the student some freedom of expression. Answering short answer questions, moreover,

involves activities such as inference making, recognition of a sequence, comparison, and establishing the main idea of a text, all of which require the relating of sentences in a text with other items which may be some distance away in the text. The disadvantage is that it requires the reader to write, and this is of some concern because it may interfere with the measurement of the intended construct.

Cloze: Although seldom used in FL tests of reading comprehension, the cloze procedure is considered by many as a valid and uniform measure of reading comprehension. Strong claims have been made for the value of the cloze procedure. It is sometimes contended, for instance, that a well-designed cloze measures not only language skills at a relatively low level (e.g. command of vocabulary, grammar, idioms), but also higher-order skills such as awareness of "intersentential relationships", global reading comprehension, etc. (see for example Chihara et al. 1979; Bachman, 1982; Bensoussan and Ramarz, 1984).

Swaffer et al., (1991) argue that the cloze procedure, while a product-oriented test, is considered as more text-based than either true-false or multiple choice answers. Recently, Bachman (1990), argues that although cloze procedures do not produce perfect tests of overall language proficiency, they do hold potential for measuring aspects of students' written grammatical competence, consisting of "knowledge of vocabulary, morphology, syntax, and phonology/graphology," and textual competence, knowledge of the cohesive and rhetorical properties of text" (pp.87-88). More recently, Oller (1992) argues that the value of pragmatic testing techniques, cloze being one of them, lies in the fact that they are based on the relatively recent theoretical linguistic terms of "text linguistics," "discourse analysis," and "pragmatics"-terms that are now popular in a growing literature, though they have a wide range of accrued meanings.

However, data that may cast doubt on the cloze as a valid assessment instrument are not lacking. Some researchers (e.g., Carroll, 1972; Lado, 1986) have questioned the

notion that successful performance on cloze test requires ability to interpret global text meanings, the implication being that cloze items are essentially sentence-bound. Other researchers have tried to define the possible limit of the range of a cloze task to 5-10 words either side of the blank (Kamil et al., 1986; Shanahan, Kamil, & Tobin, 1982). If such an estimate were to be found valid, it would mean, in effect, that cloze tasks are often insensitive to discourse constraints across sentence boundaries.

Other opponents have pointed to their failure to correspond with a test of rhetorical structure (Kintsch and Yabrough, 1982), and the fact that they are lacking in validity as a test of text-based comprehension (Farhady, 1983). In summary, cloze type techniques produce tests that can measure with some degree of accuracy, aspects of the students' written grammatical and/or textual competence. The accuracy of measurement and specific traits measured may depend on how deletions are made and the manner of students' response.

The Recall Protocol

Currently, there is almost a consensus in the L1 and L2 reading research communities that the recall of text is the best research method to obtain a performance from which we can infer what the process of comprehension is. The recall protocol is an assessment instrument in which readers are asked to read a short passage and then to write, in their native language, everything they can remember about it. Analyzing a written recall of a text by a reader is, indeed, a method which can give the researcher a fair approximation of the way the text material has been processed. Hayes (1989) has described protocol analysis as "cognitive psychology's most powerful tool for tracking psychological processes" (p. 69). Bernhardt (1991) argues that this tracking capability allows the researcher or the teacher to detect whether any lack of grammar "is interfering with the communication between the reader and text, while not focusing a reader's attention on linguistic elements in the text" (p. 200). Burton, Niles, & Waldman (1981) explain

that recall is a valid measure of reading comprehension because "under normal conditions reading is considered a semantic processing task"(p. 158). This is certainly in keeping with a cognitive approach to learning where memorization and imitation are seen as less indicative of learning than description, explanation, understanding, and elaboration.

In experimental settings in both first- and second-language research, manipulations of types of recall measure utilized have shown that free recall not only provides more valid information than any type of structured questioning but also is "the most straightforward assessment of text-reader interaction" (Johnson, 1983, p. 54). Recall, according to Bernhardt (1983) reveals "something about the organization of stored information, about some of the retrieval strategies used by readers, and reveals the method of reconstruction which [the reader] employs to encode information in a text" (Bernhardt, 1983, p.31).

Moreover, Bernhardt (1991) argues that the protocol is a valid measure of reading comprehension as it conforms to current second (L2) reading research-driven theories, such as the Constructivist Reading Model. Further, Bernhardt (1991) contends that the recall protocol "circumvents the pitfalls" (p. 28) associated with multiple choice test items because it provides no leading information or cues pertaining to passage content and requires the reader to integrate the components of the reading passage well enough to be able to recall it in a logical and coherent manner. In other words, generating recall data does not influence a reader's understanding of a text and thus "constitutes a purer measure of comprehension, uncomplicated by linguistic performance and tester interference" (p. 200).

More importantly, and in line with dynamic assessment, the recall provides considerable descriptive data about the way the subject has processed and stored the text in memory, which experimenter-directed tests rarely expose. Put differently, this procedure allows misunderstandings or gaps in comprehension to surface; a

desirable feature other measures cannot offer. Swaffer, Arens, and Byrnes (1991) note that writing the protocol in the students' native language helps reveal "how the readers' logical manipulations-their predicting, organizing, and inferencing about textual meaning-interact with their recognition of textual vocabulary and syntax" (p. 164).

In sum, recalling can add immeasurably to our understanding of readers' comprehension, "because it allows us to get a view of the quantity, quality, and organization of information gleaned during reading"(Winograd, Wixson, and Lipson, 1989, p. 123).

Based on these findings and the claims about the superiority of the recall as a comprehension measure and because of the drawbacks of objective and the so-called pragmatic tests (Bernhardt, 1991; Morrow, 1988; Ringler and Weber, 1987; Winograd , Wixson, and Lipson, 1989) suggest that teachers should make greater use of recalls.

Criticism of the Recall Protocol as a Measure of Reading Comprehension

Several L1 and L2 reading specialists and educators (Maria, 1990; Page, 1990; Swaffer, et al., 1991) have voiced criticisms of the recall protocol as a proficiency test for its inappropriateness for ESL settings, for the absence of objective weighting analyzing system, for being a time-consuming process, for its focus on holistic comprehension, and furthermore for not delineating the different processes and skills involved especially the effect of memory.

Swaffer et al. (1991) have some pragmatic objections about the procedure relating to the problematic nature of standardized grading/scoring due to the absence of a more "objective" weighting and analyzing system. More importantly, Swaffer et al., (1991) consider the measure questionable due to the absence of a "ranking system...that accounts for reader schemata" (p. 164). Instead,they suggest holistic alternatives to the recall protocol that include procedural matrices, idea maps, and story grammars which in their opinion, "reinforce instructional approaches and the

use of second language" (p. 165).

Another problem with recalling or retelling to assess comprehension is that recall is not exactly the same as comprehension. A reader may understand an idea in the text but not remember it and fail to include it in the recall. Some readers may have memory problems. Page (1990) argues that the recall of a text cannot be considered as a safe indicator of what has really been comprehended by the subject when he was reading a text. First, we are obliged to assume that many elements of textual information which are recalled have been adequately comprehended, even if we note important changes when we compare the text formulation of those elements of information with the one we read in the recall. Many cases of such changes may be considered as inferences (Fincher-Keifer, 1992; Levasseur and Page, 1989), but we can hardly consider every case or inference as an adequate comprehension of the text. We are also obliged to assume that every element of textual information which is missing in the recall has not been comprehended by the reader. Because many of those elements of information might have been comprehended but forgotten by the subject, we cannot make this assumption on safe bases.

Another problem in using recall is that some readers may have difficulty in expressing their ideas. A poor recall may be a reflection of this difficulty rather than a lack of comprehension. Another problem with recalls is that they are difficult to score. Researchers who use recall use a text analysis system to divide a text into idea units and assign those ideas to a particular level of importance. They score recalls by determining the number of text ideas they contain giving more weight to ideas with higher levels of importance. According to Maria (1990), there are two problems with this approach. First, because recalls are never in the same words of the text, deciding whether a particular idea in the recall matches an idea in the text is often difficult. Second, when researchers score recalls in the detailed way, there are always two independent scorers in order to be sure the scoring is consistent. It is unlikely that

teachers would be able to get other teachers to help them with such a time-consuming process.

Creating Recall Protocol Evaluation Instrument/Scoring Template

The means of creating a text-based instrument for evaluating recall are varied and complicated. Logical structures of text, idea units (propositions), order of presentation, and cohesion have all been utilized in a variety of studies. For an analysis of general recall that does not attempt to analyze the effects of discourse properties (redundancy, anaphora, cohesion, etc.), the most common means of creating an instrument for evaluating a free recall protocol is to weight all possible propositions in a text according to their importance (how crucial each one is for conveying the main points of the text) on a scale.

Pellegrino and Hubert (1982) note that two decades ago, free-recall, the primary paradigm and method for studying recall, gained greater prominence with Johnson's (1970) introduction of the notion of the importance of individual propositions. Meyer (1973) contends that Johnson's propositional scale constitutes a major turning point in the way recall has been evaluated. According to Kintsch & van Dijk (1978), recall measures which fail to take differential semantic importance of recalled ideas or propositions into account do not really measure both the quantity and the quality of retention. Because all evidence points to the primary value of retaining higher-level propositions and to the decreasing importance of retaining propositions as they become less and less crucial to the overall meaning of the text, propositional weighting such as that delineated by Meyer (1973) has become a generally accepted approach to evaluating recall.

Meyer(1974) recommends a scale of from one to seven. Individual protocols are searched for each proposition and awarded points commensurate with the weights of any valid propositions that are found in the reader's recall. In Meyer's hierarchical content-structure analysis (Meyer, 1975; Meyer & Freedle, 1984) an idea unit or

proposition is a meaning unit which always consists of a predicate (relation) and one or more arguments (that is, concepts connected to each other by the relation). Scoring the recall protocol involves analyzing each passage or text into a set of idea units. Each idea unit consists of a single clause (main or subordinate, including adverbial and relative clauses). Each infinitival construction, gerundive, nominalized verb phrase, is also identified as a separate idea unit. In addition, optional/or heavy prepositional phrases are also designated as separate idea units.

Idea units are organized into a hierarchy (Meyer, 1975; Meyer & Freedle, 1984). Each idea unit is determined to be a top-, high-, mid-, or low-level idea unit, according to the following criteria:

1. Top-Level: Represents the main ideas being compared or contrasted or the main ideas being collectively described.
2. High-level: Represents major ideas or main topics in the passage.
3. Mid-Level: Represents minor ideas or subtopics in the passage.
4. Low-Level: Represents minor detail in the passage.

According to the Johnson (1970) analysis system, a reading passage is divided into pausal units during a normally paced oral reading. Each pausal unit or proposition is weighted on a scale of one to four depending on its importance to the passage content, one being least important and four most important. The weighting usually reflects the mean of the ratings given to each proposition by proficient readers.

Once the proportions are weighted, a scoring template can be developed and followed when scoring readers' recall protocols. According to this procedure, the total score on a recall is the sum of the scores on each proposition. Propositions, therefore, are treated as discrete-point items, as are multiple choice test items

Rationale for a Qualitative Analysis of the Recall

The different quantitative scoring methods are not helpful in terms of indicating what parts of the text are particularly problematic for students. Quantitative scoring

systems lack provisions for determining the sort of specific errors students make and how those errors impede their comprehension. From this perspective, Berkemeyer (1989) argues that the scoring instrument should focus not on what students do or do not recall instead on what they attempt to recall or to integrate into their protocol but fail to do so correctly. One could also argue that indications of misinterpretations or restructurings in the recalls that reflect a lack of comprehension are also lacking as well as indications of whether the reader uses the structure of the text to structure his or her recall. It is only through qualitative analysis that the teacher or researcher can begin to discover what is impairing students' comprehension processing and why. This information may ultimately be of more value to the classroom teacher, because it may suggest ways to adjust instruction in order to promote better reading comprehension.

To arrive at this kind of deeper information, several alternative analysis procedures of a more qualitative nature have been suggested both in L1 and L2 contexts. The trade-off is that some of the objectivity obtained by means of the binary scoring system will be lost, but at the same time important and deeper insights into the processes and constructive activities involved in reading will be gained as well as indications of the different types of troublesome textual features (Berkemeyer, 1989).

A useful tool for such analysis is Bernhardts' model of L2 text comprehension. The model elucidates in a direct way the kinds of errors students are making. As a qualitative model, it emphasizes not so much the product but the process of comprehension. In so doing, it reveals the "patterns of intrusions, distortions, and omissions which provide much valuable information for understanding the comprehension process. Unlike the Meyer's-based scoring system, Bernhardt's model focuses on the connected interactions between various textual features and influences external to the text.

Rationale for Using Constructive Activity Scale for Scoring the Recall Protocol

Current quantitative scoring methods of the recall protocol reward or penalize the reader for the presence or absence of previously determined and weighted propositions in his or her recalled text. Current rating systems, however, ignore and do not reward attempts aimed at paraphrasing or summarizing information-two important skills indicative of deep-level or active processing. They also ignore the reader's relevant elaborations aimed at assimilating and subsequently integrating text-based information into his or her cognitive structure.

The distinction between a less active and more active pursuit of knowledge has become an important theme in instructional psychology (Brown, Bransford, Ferrera, and Campione, 1983; Resnick, 1989). Such concepts as explanation-driven learning (Brown and Kane, 1988) meaning imposition (Resnick, 1987), mindfulness (Salomon and Globerson, 1987) and intentional learning (Bereiter and Scardamalia, 1989) are among those used to characterize active learning, in which more extensive or deeper constructive activity occurs.

In addition, studies in text processing (Einstein, McDaniel, Owen, & Cote, 1990) suggest that for a processing activity to be effective in reading, readers need to encode both relational and individual-item information or the full set of information and the different elements which form this full set in the text and that "different types of materials and processing activities encourage encoding of different types of information" (Einstein, McDaniel, Owen, & Cote, 1990, P.570).

The cornerstone of this framework is the assumption that the two types of information are essential for the production of good free recall (Einstein and Hunt, 1980; Hunt and Einstein, 1981; Hunt and Marschark, 1987). Eisentien et al. (1990) define individual-item information as that specific to propositions or individual concepts or within the stimulus material. Relational information, on the other hand, represents the integration or organization of the individual propositions

within the text.

The Constructive Activity Scale for Rating the Recall Protocol

This author proposes a scale that represents four levels (two low and two high) of cognitive constructive activity involved in comprehension and learning from text. The two low-level or less active constructive activities involve restating/retelling text information or making inferences based on the surface features of the text. The two high-level or more active constructive activities involve problem solving activities exemplified in carrying out meaningful inferencing, problem solving, information reconciliation, assimilation, and integration.

The proposed scale is quantitative in nature but has provision for analyzing readers' recalls qualitatively. It is adapted from Chen, Burtis, Scardamalia, and Bereirter's (1992) scale for cognitive constructive activity in learning from text. The four levels are:1- Prepropositional/Fragmented Associations, 2-Knowledge/Details Retelling, 3-Assimilation, and 4-Problem Solving and Integration. The examples used in this paper are taken from recalls of beginning American students learning Arabic as a foreign language (AFL).

Level 1: Prepropositional/Fragmented Associations: A rating of 1 is assigned to responses that depend on isolated words or fragmented phrases and do not show an understanding of the text at a propositional level. Overextended inferences, associations of irrelevant personal knowledge, comments, and responses involving associative reactions to words or brief fragments that do not deal with what the text says about a particular vocabulary word are assigned 1. For example, the text statement "He used to watch a film once a week at the 'Radio City" theater in downtown Cairo" was recalled by one student as "When he was studying in Cairo, he would listen to a program from Radio City." Notice that this reader has reacted to "Radio City" which he interpreted as radio station and accordingly recalled "listen to a program" although the verb "watched" and the noun "theater" are clear in the text

statement.

Interrogative fragments or responses that question the meaning of isolated words out of context are also assigned 1. All these types of responses show minimal, low-level, or less active constructive activity on the part of the reader. This level is characterized by item-by-item approach to text processing.

Level 2: Knowledge/Details Retelling: It is argued that cases in which the reader matches the surface features of text propositions with what he or she knows set off a

process of knowledge retelling. For example, the text statement "Samir came from Syria and studied at Georgetown Universtiy" was recalled by a reader who is clearly familiar with the geography of the Middle East as "The person is from Syria. He is studying at the university. The person is from Damascus, the capital of Syria. Damascus is an old city and has many people." Notice that the word "Syria" has prompted the reader to start the process telling information that he knows but is not part of the text statement.

This association of knowledge normally involves no clarification or elaboration of the text meaning and does not reflect how things have taken place. A rating of 2 is assigned to verbatim or near-verbatim paraphrases of the text (detail recalling) and knowledge recalling. Whereas associations at Level 1 are cued by isolated words, Level 2 associations generally involve the association of topically related personal knowledge cued by a text proposition. Although the text is processed at a deeper level, that level is still shallow. Level 2 propositions normally lack integration of text information with personal knowledge and are characterized by dominance of either text-based or knowledge-based information (mostly text-based information).

Level 3: Assimilation: Propositions that show evidence of text-based representation of information are assigned Level 3. Propositions that involve paraphrasing and adding simple relevant elaborations provide evidence of text comprehension and

show grasp of what the text says but fail to reconcile the text message with the accepted or more dominant notions in the field. In other words, propositions at Level 3 suggest reader's ability to construct text representation but with no attempts made to use the assimilated knowledge to transform or reformulate his or her cognitive structure.

For example, a reader paraphrased a large part of a text--a cover letter sent by a doctoral student to the chairperson of the English Department at the University of Kuwait--by the following sentence " He is writing to Kuwait University looking for a position in the coming year in the Department of Foreign Languages." This reader is showing comprehension of textual information as is evidenced in his brief recall.

Level 4: Problem Solving: Propositions that reflect attempts to reconcile and integrate text information into the reader's existing knowledge structure are scored as level 4. Such attempts are indicative of a high-level problem solving constructive activity in which inconsistencies or discrepancies between text-based and knowledge-based information are resolved by means of hypothesis generation. For example, one of the statements in a text on Egypt reads as follows: "Egypt depends on the waters of the Nile River. The Roman historian Herodotus described Egypt as "the gift of the Nile." A student recalled this statement as follows" Egypt depends on the waters of the Nile. Herodutus (by the way, a Greek historian, not Roman) described Egypt as the gift of the Nile." This reader has identified and resolved discrepancy between text information and his personal knowledge. This response may be referred to as "evaluative response." At Level 4, readers use multiplicity of relations to attend to new information with the resultant of forming new connections. Attempts made to use knowledge-based information to explain text statements are also rewarded as well as attempts to relate and integrate earlier statements in the text to the current statement (relational information). In a different text, a statement reads as follows: "The student came to the U.S. to study at

the Ohio State University." Earlier in the text, we are told that the student comes from a poor family. The reader recalled the following: "The student is relating how he came to study at the Ohio State University. He must have received a scholarship." In attempting to make sense of text information, this reader related earlier information in the text and formulated a hypothesis about the student's ability to study at a U.S. university despite the financial status of his family.

Analysis of the relations between the two facets of reading--knowledge-driven and text-driven--reflects the construction of a situation model--as opposed to text representations that characterize Level 3 on the constructive activity scale--and opens the way to new understanding.

In sum, recalling or retelling important and relevant information directly stated or inferred from text indicates the reader's comprehension of textual information. Connecting and integrating text information and reader's background knowledge, summarizing statements or making generalizations based on text information, reacting to text information all indicate metacognitive awareness and strategy use. Appropriate use of language in the recall, awareness of the structure of the text, and the ability to organize the recall in an acceptable format indicate facility with language.

The present rating scale is an attempt to account for both knowledge- and text-driven facets of reading comprehension as well as the interaction among the reader, text, and author. The failure to do so defeats one of the major purposes of reading, namely communicating meaning to the reader regardless of how the reader's role is conceptualized (mere recipient of information, problem solver, or an active participant) in the assignment and construction of meaning.

The proposed rating scale, moreover, rewards the author of the recall protocol for his or her attempts to construct text representation or situation models of the text. After all, learning from the text involves more than the comprehension of

additional information in the text; it involves active construction of new understanding and new knowledge (Chen, et al., 1992).

In addition to that, the proposed system and by virtue of providing qualitative information about text processing, serves an important pedagogical purpose by enhancing the process of dynamic assessment. Misconceptualizations, misunderstandings, gaps, distortions, and elaborations in the reader's protocol provide great insights into the teacher or the researcher about text processing strategies. Detecting and delineating such information can help reveal problems that impede comprehension.

We are beginning to explore how current theories advanced from cognitive psychology and information-processing perspectives can be best utilized in reshaping our pedagogical and assessment practices. The proposed scale for calibrating and rating the recall protocol may be a potentially important tool for assessing reading comprehension. This is consonant with the call for the need of grounding comprehension measures in models and theories of learning. There is a need, however, to empirically demonstrate the reliability, validity, and usefulness of the proposed scale to understand more fully the process of learning and reading comprehension.

BIBLIOGRAPHY

Afflerbach, P. 1990. "The Influence of Prior Knowledge on Expert Readers' Main Idea Construction Strategies." *Reading ResearchQuarterly* 25(1): 31-46.

..... . 1986. "The Influence of Prior Knowledge on Expert Readers' Importance Assignment Processes," pp. 30-40 in J. Niles, and R. Lalik eds., *Solving Problems in Literacy: Thirty-FifthYearbook of the NRC*,. Rochester, NY: National Reading Conference.

Alexander, P., and Judy, J. 1988. "The Interaction of Domain-Specific Knowledge and Strategic Knowledge in Academic Performance." *Review of Educational Research* 58: 375-404.

Anderson, N., Bachman, L., Perkins, K., and Cohen, A. 1991. "An Exploratory Study in the Construct Validity of a Reading Comprehension Test: Triangulation of Data Sources. "*Language Testing*, 8(1): 41-66.

Anderson, R., Reynolds, R., Schallert, D., and Goetz, E. 1977. "Frameworks for Comprehending Discourse." *American Educational Research Journal*14: 367-381.

Anderson, R. C., and Pearson, P. 1984. "A Schemata-Theoretic View of Basic Pocesses in Reading Comprehension," pp. 255-292 in P. Pearson, R. Barr, M. Kamil, and P. Mosenthal, eds.,*The Handbook of Reading Research*. New York: Longman.

Ashcraft, M. H. 1989. *Human Memory and Cognition*. New York: Scott, Foreman, and Company.

Bachman, L. 1982. "The Trait Structure of Cloze Test Scores."*TESOL Quarterly* 16: 61-70.

---------- . 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bensoussan, M., and Ramraz, R. 1984b. "Testing EFL Reading Comprehension Using a Multiple-Choice Rational Cloze."*Modern Language Journal* 68: 230-39.

Bereiter, C., and Scardamalia, M. 1989. "Intentional Learning as a Goal of Instruction," pp. 12-26 in L. B. Resnick, ed., *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bereiter, C., Butris, P., and Scardamalia, M. 1988. "Cognitive Operations in Constructing Main Points in Written Composition." *Journal of Memory and Language* 7: 261-278.

Berkemeyer, V. B. 1989. "Qualitative Analysis of Immediate Recall Protocol Data: Some Classroom Implications." *Die Unterrichtspraxis* 22: 131-137.

Bernhardt, E. 1991. *Reading Development in a Second Language:Theoretical, Empirical, and Classroom Perspectives.* Norwood, NJ: Ablex Publishing Corp.

..... . 1984. "Text processing Strategies of Native, Non-native Experienced, and Non-native Inexperienced Readers of German: Findings and Implications for the Instruction of German as a Foreign Language." Unpublished doctoral dissertation, The University of Minnesota, Minneapolis, MN.

..... . 1983. "Testing Foreign Language Reading Comprehension: The Immediate Recall Protocol." *Die Unterrichtspraxis* 16(1): 27-33.

Brown, A. L., and Kane, M. J. (1988). "Preschool Children Can Learn to Transfer: Learning to Learn and Learning from Example."*Cognitive Psychology* 20: 493-523.

Brown, A. L., Bransford, J. D., Ferrera, R. A., and Campione, J. C. 1983. "Learning, Remembering, and Understanding," pp. 77-166 in J. H. Flavell and E. M. Markman, eds., *Handbook of child psychology:Vol. 3. Cognitive development* . New York: John Wiley & Sons.

Burton, J. K., Niles, J. A., and Wildman, T. M. 1981. "Levels of Processing Effects on the Immediate and Delayed Recall of Prose." *Journal of Reading Behavior* 13: 157-164.

Campione, J. C., and Brown, A. L. 1982. *"Dynamic Assessment: One Approach and Some Initial Data."* Technological Report No. 361. Urbana, IL: Center for the Study of Reading.

Carroll, J. B. 1976. "Psychometric Tests as Cognitive Tasks: A New "Structure of Intellect," pp. 102-121 in L. B. Resnick, ed., *The Nature of Intelligence.* Hillsdale, NJ: Erlbaum.

Casanave, C. P. (1988). "Comprehension Monitoring in ESL Reading: A Neglected Essential."*TESOL Quarterly* 22(2): 283-302.

Chan, K. K.,, P. J., Burtis, P. J., Scardamalia, M., & Bereiter, C. 1992. "Constructive Activity in Learning from Text."*American Educational Research Journal* 29: 97-118.

Chihara, T., Oller, J., Weaver, K., & Chavez-Oller, M. 1977. "Are cloze Items Sensitive to Constraints Across Sentences?" *Language Testing* 27(1): 63-69.

Cronbach, L. J. 1984. *Essentials of Psychological Testing*. NY: Harper and Row.

Cross, D. R., and Paris, S. G. 1987. "Assessment of Reading Comprehension: Matching Test Purposes and Test Properties." *Educational Psychologist* 22: 313-322.

Daneman, M., and Tardif, T. 1988. "Working Memory and Reading Skill Re-examined," pp. 491-508 in M. Colheart, ed., *Attention and Performance* . London: Erlbaum.

Dewitz, P., Carr, E. M., & Patberg, J. P. 1987. "Effects of Inference Training on Comprehension and Comprehension Monitoring." *Reading Research Quarterly* 22: 99-119.

Doyle, W. 1983. "Academic Work." *Review of Educational Research*53: 14-21.

Einstein G., McDaniel, M., Owen, P., and Cote, N. 1990. "Encoding and Recall of Texts: The Importance of Material Appropriate Processing." *Journal of Memory and Language* 29: 566-581.

Einstein, G. O., and Hunt, R. R. 1980. "Levels of Processing and Organization: Additive Effects of Individual-Item and Relational Processing." *Journal of Experimental Psychology: Human Learning and Memory*6: 588-598.

Farhady, H. 1983. "New Directions in ESL Proficiency Testing," pp. 253-269 in J. Oller Ed., *Issues in Language Testing Research*. Rowley, MA: Newbury House.

Farr, R., Pritchard, R., and Smitten, B. 1990. "A Description of What Happens When an Examinee Takes a Multiple-Choice Reading Comprehension Test." *Journal of Educational Measurement*27:209-226.

Fincher-Kiefer, R. 1992. "The Role of Prior Knowledge in Inferential Processing." *Journal of Research in Reading* 5: 12-27.

Hansen, J. 1981. "The Effect of Inference Training and Practice on Young Children's Reading Comprehension." *Reading Research Quarterly*16: 391-417.

Hansen, J., and Pearson, P. 1983. "An Instructional Study: Improving the Inferential Comprehension of Good and Poor Fourth-Grade Readers." *Journal of Educational Psychology*75: 821-29.

Hayes, J. R. 1989. *The Complete Problem Solver*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Henning, G. 1987. *A guide to Language Testing: Development, Evaluation, and*

Research. Cambridge, MA: Newbury House Publishers.

Hunt, R. R., and Einstein, G. O. 1981. "Relational and Item-Specific Information in Memory." *Journal of Verbal Learning and Verbal Behavior* 20: 497-514.

Johnson, R. 1970. "Recall of Prose as a Function of the Structural Importance of Linguistic Units." *Journal of Verbal Learning and Verbal Behavior* 9: 12-20.

Johnston, P. H. 1983. *Reading Comprehension Assessment:A cognitive Basis.* Newark, DE: International Reading association.

Kamil, M. L., Smith-Burke, M., & Rodriguez-Brown, F. 1986. "The Sensitivity of Cloze to Intersentential Integration of Information in Spanish Bilingual Populations,"pp. 334-338 in J. Niles and R. Lalik, eds., *Thirty-Fifth Yearbook of the National Reading Conference.* Rochester, NY: National Reading Conference.

Kintsch, W. 1974.*The Representation of Meaning in Memory*.Hillsdale, NJ: Lawrence Erlbaum.

Kintsch, W., and van Dijk, T. A. 1978. "Toward a Model of Text Comprehension and Production." *Psychological Review* 85: 363-394.

Kintsch, W., and Yarbough, J. C. 1982. "Role of Rhetorical Structure in Text Comprehension." *Journal of Educational Psychology*74: 828-834.

Klein-Braley, C. 1991. "Ask a Stupid Question...: Testing Language Proficiency in the Context of Research Studies," pp. 73-94 in K. de Bot, R. Ginsberg, and C. Kramsch, eds., *Foreign Language Research in Cross-Cultural Perspective.* Amsterdam: John Benjamin Publishing Co.

Klein-Braley, C. 1983. "A Cloze is a Question," pp. 23-46 in J. Oller ed., *Issues in Language Testing Research.* Rowley, MA: Newbury House.

---------- . 1985. "A Cloze-Up on the C-Test." *Language Testing* 2: 76-104.

Klein-Braley, C., and Raatz, U. 1984. "A Survey of Research on the C-Test." *Language Learning* 1: 134-146.

Lado, R. 1986. "Analysis of Native Speaker Performance on a Cloze Test." *Language Testing* 3: 130-146.

Levasseur, D, and Page, M. 1989. "Comprehension of Sentences and of Inter-sentential Relations by 11 to 15 Year Old Pupils," pp. 133-147 in J. H. De Jong and D. K. Stevenson, eds., *Individualizing the Assessment of Language Abilities.* Clevdon: Multilingual Matters.

Maria, K. 1990. *Reading Comprehension Instruction: Issues and Strategies.* Parkton, MD: York Press.

Meyer, B. J. 1974. *"Effects of Position of Information in a Passage's Organizational structure on Recall."* Paper Presented at the Annual Meeting of the American Educational Research Association, Chicago, IL: (ERIC ED 090 525).

Meyer, B. J. 1975. *The Organization of Prose and its Effects on Memory.* New York: North-Holland.

Meyer, B. J., and Freedle, R. O. 1984. "Effects of Discourse Type on Recall." *American Educational Research Journal*21: 121-143.

Morrow, K. 1979. "Communicative Language Testing: Revolution or Evolution?" pp. 143-157 in C. Brumfit and K. Johnson, eds.,*The Communicative Approach to Language Teaching.* Oxford: Oxford University Press.

Nivo, N. 1989. "Test-Taking Strategies on a Multiple-Choice Test of Reading Comprehension." *Language Testing* 6: 199-215.

Oller, J. Jr. 1992. "Foreign Language Testing, Part 2: Its Depth." *ADFL Bulletin* 22(4): 5-13.

Oller, J. W., Jr. 1979. *Language Tests at School: A Pragmatic Approach.* London: Longman.

Oscarson, M. 1991. "Item Response Theory and Reduced Redundancy Techniques: Some Notes on Research Developments in Language Testing," pp. 95-111 in K. de Bot, R. Ginsberg, and C. Kramsch, eds., *Foreign Language Research in Cross-Cultural Perspective.* Amsterdam: John Benjamin Publishing Co.

Page, M. 1990. "Methodological Issues in Testing Comprehension of Texts," pp.113-28 in M. Spoelders, ed., *Literacy Acquisition.* Tubingen: J. Van In.

Pellegrino, J. W., and Hubert, L. J. 1982. "The Analysis of Organization and Structure in Free Recall," pp. 129-172 in C. R. Puff, ed.,*Handbook of Research Methods in Human Memory and Cognition.* New York: Academic Press.

Pritchard, R. 1990. "The Effects of Cultural Schemata on Reading Processing Strategies." *Reading Research Quarterly*25: 273-295.

Pyrczak, F. 1975. "Passage Dependence of Reading Comprehension Questions: Examples." *Journal of Reading* 18: 308-311.

Resnick, L. B. ed. 1989. *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser.*.Hillsdale, NJ: Lawrence Erlbaum Associates.

..... .1987. *Education and Learning to Think.* Washington, D.C: National Academy Press.

Ringler, L. H., and Weber, C. K. 1984. *A Language-Thinking Approach to Reading: Diagnosis and Teaching.* New York: Harcourt, Bruce, and Javanovich.

Rothkopf, E. Z. 1972. "The Concept of Mathemagenic Activities." *Review of Educational Research* 40: 325-336.

Rumelhart, D. E. 1980. "Schemata: The Building Blocks of Cognition," pp. 33-58 in B. Spiro, B. Bruce, and W. Brewer, eds., *Theoretical Issues in Reading Comprehension* . Hillsdale, NJ: Lawrence Erlbaum.

Rumelhart, D. E. 1977. "Toward an Interactive Model of Reading," pp. 573-603 in S. Dornic, ed., *Attention to Performance* . New York: Academic Press.

Rumelhart, D. E., and Ortony, A. 1977. "The Representation of Knowledge in Memory," pp. 99-135 in R. Anderson, R. Spiro, and W. Montague, eds., *Schooling and the Acquisition of Knowledge* . Hillsdale, NJ: Erlbaum.

Salomon, G., and Globerson, T. 1987. "Skill is Not Enough: The Role of Mindfulness in Learning and Transfer."*International Journal of Research in Education* 11: 623-638.

Shanahan, T., Kamil, M., and Tobin, A. 1982. "Cloze as a Measure of Inter-sentential Comprehension."*Reading Research Quarterly*17: 229-225.

Spiro, R. 1980. "Constructive Processes in Prose Comprehension and Recall," pp. 245-278 in R. Spiro, ed.,*Theoretical Issues in Reading Comprehension.* Hillsdale, NJ: Erlbaum.

Swaffer, J., Arens, C., and Byrnes, H. 1991. *Reading for Meaning: An Integrated Approach to Language Learning.* Englewood Cliffs, NJ: Prentice-Hall Inc.

Thorndike, R. 1973. *Reading Comprehension Education in Fifteen Countries.* NY: Wiley.

---------- . 1917a. "Reading as Reasoning: A Study of Mistakes in Paragraph Reading." *Journal of Educational Psychology* 8: 323-332.

---------- . 1906. *Principles of Teaching.* New York: Seiler.

Valencia, S. W., and Pearson, P. D. 1988. "Principles for Classroom Comprehension Assessment." *Remedial and Special Education* 9: 26-35.

Valencia, S. W., McGinley, W., and Pearson, P. D. 1991. "Assessing Reading and

Writing: Building a More Complete Picture," in G. Duffy, ed.,*Reading in the Middle School*. New York: International Reading Association.

van Dijk, T. A. 1977. "Semantic Macro-Structures and Knowledge Frames in Discourse Comprehension," pp. 3-32 in M. A. Just, and P. A. Carpenter, eds., *Cognitive Processes in Comprehension*. Hillsdale, NJ: Lawrence Erlbaum.

..... . 1980. *Macrostructures*. Hillsdale, NJ: Erlbaum.

van Dijk, T., and Kintsch, W. 1983. *Strategies for Discourse Comprehension*. New York: Academic Press.

Venezky, R. L. 1984. "The History of Reading Research," in P. D.Pearson, ed., *Handbook of Reading Research*. New York: Longman.

Wang, M. C. 1987."The wedding of Instruction and Assessment in the Classroom." Paper presented at the Educational Testing Service Invitational Conference, Princeton, NJ: Educational Testing Service.

Weir, C. J. 1990. *Communicative Language Testing*. NY: Prentice Hall.

Winograd, P. N., Wixson, K. K., and Lipson, M.Y. 1989. *Improving Basal Reading Instruction*. NY: Teachers College Press.

Wood, N. Y. 1989. "Reading Tests and Reading Assessment. *Journal of Developmental Education*13: 14-18.

Woods, D. 1988. "Teachers' Interpretation of Language Teaching Materials." Paper Presented at the 1988 RELC Seminar, 11-15 April, 1988, Singapore.